






## Article

# Adaptive Online-Learning Volt-Var Control for Smart Inverters Using Deep Reinforcement Learning

Kirstin Beyer , Robert Beckmann , Stefan Geißendörfer , Karsten von Maydell  and Carsten Agert 

German Aerospace Center (DLR)—Institute for Networked Energy Systems, Carl-von-Ossietzky-Straße 15, 26129 Oldenburg, Germany; robert.beckmann@dlr.de (R.B.); Stefan.Geissendoerfer@dlr.de (S.G.); Karsten.Maydell@dlr.de (K.v.M.); Carsten.Agert@dlr.de (G.A.)

\* Correspondence: kirstin.beyer@dlr.de

**Abstract:** The increasing penetration of the power grid with renewable distributed generation causes significant voltage fluctuations. Providing reactive power helps balancing the voltage in the grid. This paper proposes a novel adaptive volt-var control algorithm on the basis of deep reinforcement learning. The learning agent is an online-learning deep deterministic policy gradient that is applicable under real-time conditions in smart inverters for reactive power management. The algorithm only uses input data from the grid connection point of the inverter itself; thus, no additional communication devices are needed and it can be applied individually to any inverter in the grid. The proposed volt-var control is successfully simulated at various grid connection points in a 21-bus low-voltage distribution test feeder. The resulting voltage behavior is analyzed and a systematic voltage reduction is observed both in a static grid environment and a dynamic environment. The proposed algorithm enables flexible adaption to changing environments through continuous exploration during the learning process and, thus, contributes to a decentralized, automated voltage control in future power grids.



**Citation:** Beyer, K.; Beckmann, R.; Geißendörfer, S.; von Maydell, K.; Agert, C. Adaptive Online-Learning Volt-Var Control for Smart Inverters Using Deep Reinforcement Learning. *Energies* **2021**, *14*, 1991. <https://doi.org/10.3390/en14071991>

Received: 19 February 2021

Accepted: 29 March 2021

Published: 3 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** deep reinforcement learning; low-voltage grid; reactive power; smart inverter; voltage control; volt-var-optimization

## 1. Introduction

The proceeding decentralization of the power system due to higher penetration with renewable distributed generation (DG) can cause voltage problems in the distribution grid, since bidirectional power flows increase the risk of voltage violations [1,2]. Furthermore, the fluctuating character of renewable energy feed-in causes rapid changes in the power fluxes and affects the voltage behavior significantly [3,4]. To overcome these problems, reactive power injection or absorption can be used to compensate voltage changes [5]. Thus, volt-var control (VVC) is essential for stable grid operation [6].

Various volt-var control approaches are currently under research, such as automatic voltage regulators, switchable capacitors, distribution static compensators (DSTATCOM), and smart inverters (SI) [7–10]. This work focuses on VVC in smart inverters to combine DG and voltage control efficiently and provide a fast response to voltage changes. Various volt-var strategies for SI have been investigated recently [10]. Reactive power can either be provided with a fixed power factor, with fixed reactive power  $Q$  or voltage-dependent reactive power  $Q(U)$  [11,12]. Since the cable length and the grid topology influence the local voltage behavior, the reactive power demand at each grid connection point (GCP) is individual [13,14]. Thus, regarding the voltage-dependent approach, research is being conducted on optimizing the  $Q(U)$  feed-in [15,16].

The recent success of deep reinforcement learning (DRL) in many fields, including games [17,18] and robotics [19,20], has also attracted attention in power and energy applications [21,22]. Several studies focus on applying DRL methods to power grid operations in order to provide intelligent control algorithms [23–25]. In Reference [26,27], an optimized

coordination of various voltage regulating devices was realized through DRL. Several works suggest DRL algorithms for multi-agent smart inverter coordination [28–31] requiring knowledge from other grid nodes. Especially the deep deterministic policy gradient (DDPG) [32] is highly promising in the field of control algorithms. In Reference [33], a two-stage deep reinforcement learning method for inverter-based volt-var control is presented consisting of an offline and online stage in the learning process.

In contrast to most of these studies that use centralized methods with the need for measurement and communication devices within the grid, in this paper, a fully decentralized DRL volt-var control is demonstrated. Within the DRL framework, a DDPG learning agent is used, which only gets input data from its own connection point to the grid and does not require any previous knowledge of the grid. Hence, the proposed method can be applied to single inverters individually and is able to regulate the local voltage without any further measurement or communication equipment. This strongly reduces the computational costs and the amount of data and, thus, increases the potential for real-time applications. Additionally, the individual implementation allows a progressive and demand-orientated upgrading of the inverters in the grid without complex guidance through the distribution grid operator (DGO). Furthermore, the learning process is fully realized as an online-learning, allowing ongoing exploration. This enables the control algorithm to continually adapt to fluctuating power flows in the short term and changing grid contributors in the long term.

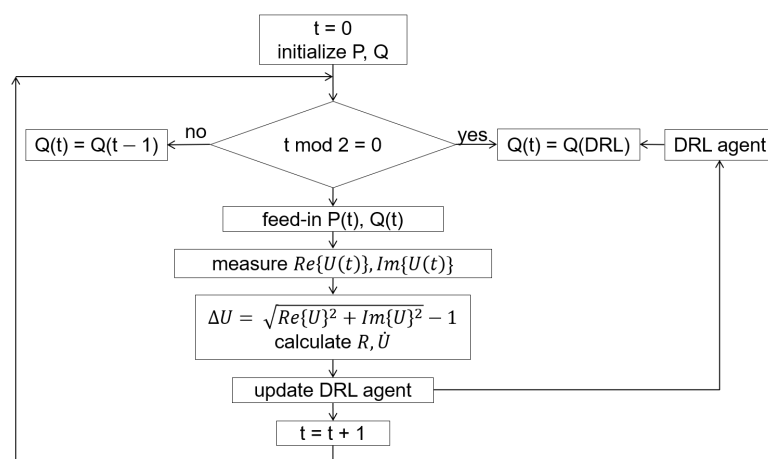
The main contributions of this paper are summarized as listed below:

- No need for additional equipment: this saves installation effort and costs.
- Reduction of data flows: with a self-learning individual volt-var control there is no need for data exchange with other grid actors.
- Individual application at the point of demand: this enables DGOs to progressively adapt the distribution grid to higher shares of DG feed-in.
- Flexible adaptation to changing environments through online-learning: the ongoing exploration in the learning process allows a continuous adaption to the actual reactive power demand.

The remainder of the paper is organized as follows: Section 2 presents the proposed deep reinforcement learning method, including the formulation of the DRL agent parameters and the reward function. Subsequently, in Section 3, the simulation test feeder is illustrated and analyzed regarding its reactive power demand at each connection point. Accordingly, the proposed DRL method is applied at various nodes in the test feeder, and the static and dynamic grid behavior is investigated in Section 4.

## 2. Proposed DRL Volt-Var Control Algorithm

The proposed voltage control method is based on a reinforcement learning framework consisting of a training environment and a deep learning agent. The learning agent used in this algorithm is a deep deterministic policy gradient (DDPG). The DDPG agent is a model-free actor-critic deep reinforcement learning agent for continuous action spaces. The algorithm was developed in 2015 by Lillicrap et al. [32] based on the deterministic policy gradient (Silver et al. [34]). It uses an actor and critic architecture with deep neural networks as function approximators. The learning agent interacts with its environment by observing the environment and receiving rewards for the performed actions. Together with the observed values, this reward is used to update the neural networks and, thus, influences the output values. In the following, this learning method is applied to volt-var-control in inverters. The algorithm is developed in such a way that only input data from the grid connection point of the inverter is used. Therefore, the observed data (input) is limited to the voltage values at the grid connection point, while the performed action (output) of the learning agent is the reactive power output. The flowchart of the proposed algorithm is shown in Figure 1.



**Figure 1.** Flowchart of the proposed deep reinforcement learning volt-var control.

To initialize the algorithm, the active and reactive power at the grid connection point are set. Every second step, the reactive power is calculated by the DRL agent and then kept constant for two steps in order to prevent an oscillating behavior. Afterwards, the active and reactive power are fed in by the inverter to the electric grid and the resulting voltage values  $\text{Re}\{U\}$  and  $\text{Im}\{U\}$  (real and imaginary part of the voltage) at the grid connection point are measured. Subsequently, the voltage deviation  $\Delta U$  between the measured voltage and the nominal voltage of 1 pu, as well as the temporal derivative  $\dot{U}$  and the reward  $R$ , are calculated. For updating the DRL agent, the values of  $\Delta U$ ,  $\text{Re}\{U\}$ ,  $\text{Im}\{U\}$  and  $\dot{U}$  are used as observation values (input). After updating the DRL agent, the new reactive power output  $Q(t)$  is calculated as the action value. For calculating the reward  $R$ , a specific reward function was developed. The reward function is essential for reinforcement learning algorithms because it defines the behavior of the DRL agent. For the considered application, the voltage is supposed to be regulated to 1 pu; therefore, any voltage deviations result in negative rewards. The developed reward function weights the absolute voltage deviation with the factor 1000 and combines this with an additional term that rewards long lasting voltage stability:

$$R = -1000 |\Delta U| - \frac{1}{1+b} \quad (1)$$

$$\text{with } b = \begin{cases} 0, & |\Delta U| \geq a \\ b+1, & |\Delta U| < a \end{cases}$$

where  $a$  describes the admissible absolute voltage deviation in pu.

This specific reward function was developed in order to prevent an oscillating voltage behavior. Taking into account the parameter  $b$ , the function rewards successive compliance of  $\Delta U$  with the desired interval  $[-a, a]$ . The longer the voltage is kept within the interval  $1 \pm a$ , the bigger the value  $b$  gets; thus, the term  $\frac{1}{1+b}$  tends to zero, leading to a higher total reward.

The proposed algorithm was implemented in Python with the help of OpenAI Gym [35] and keras-rl [36]. Optimized parameters for the DRL agent were set as listed in Table 1.

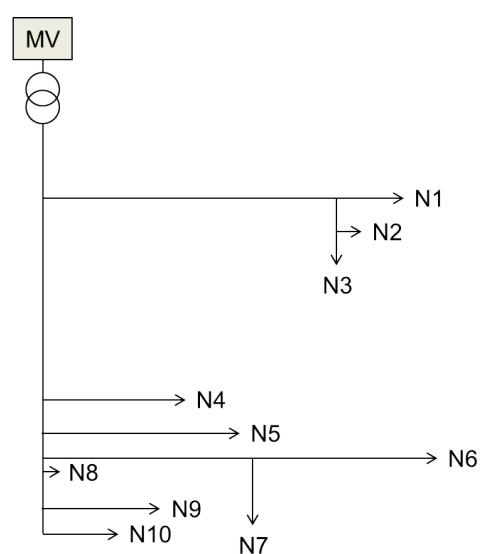
**Table 1.** Deep reinforcement learning (DRL) agent parameters.

Parameter	Value
Actor	5 layer à 32 nodes
Critic	6 layer à 64 nodes
activation function	ReLU, output layer: linear
learning rate $\alpha$	0.001
target model update $\tau$	0.001
discount factor $\gamma$	1
random process	Ohrnstein-Uhlenbeck process ( $\theta = 0.01, \sigma = 0.01$ )
warmup steps	10,000
memory	100,000

### 3. Simulation Framework

#### 3.1. 21-Bus Test Feeder

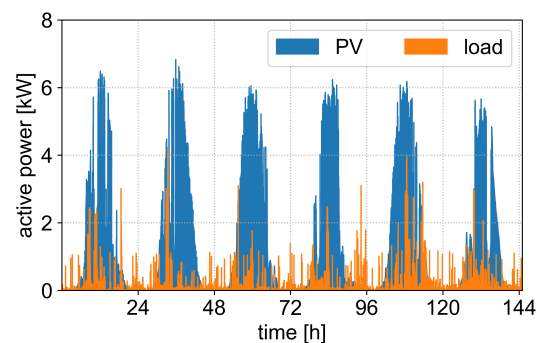
The proposed algorithm was tested in the feeder shown in Figure 2. The test feeder was developed in the study ‘Merit Order Netzausbau 2030’ (MONA) and is a three-phase 21-bus system [37]. It can be considered a European low voltage distribution grid with ten residential households at a voltage level of 400 V and a frequency of 50 Hz. The test feeder was modeled in MATLAB Simulink and exported to the training environment in Python as a functional mock-up unit (fmu-file) with FMKit [38].

**Figure 2.** Schematic representation of the ‘Merit Order Netzausbau 2030’ (MONA) 21-bus test feeder.

The simulation scenario combines distributed generation through photovoltaic (PV) at every household (N1 to N10) together with individual active and reactive loads from the households. For this, load profiles by Tjaden et al. [39] were used that provide three-phase values for active and reactive power at every second. The first ten profiles were utilized. For the PV feed-in, a normalized PV profile [40] was multiplied by a fixed factor for every household (see Table 2) to model different PV sizes. These factors were chosen randomly around an average PV size of 5 kWp for residential PV installations. For all following simulations, the three phases are assumed symmetric; thus, only the results of phase A are presented. Figure 3 shows the load and PV profile at node N10 exemplarily.

**Table 2.** Load and photovoltaic (PV) profile numbers and input values for the dynamic and static case studies. For the static case, the data from 21 June, 11.06 h 40 s was used.

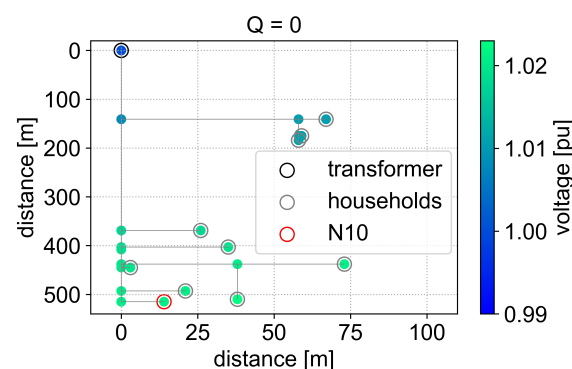
Node N	Load Profile No.	PV Factor/kWp	$P_{load}/W$ (Static)	$P_{PV}/W$ (Static)
N1	1	6.5	53	1932
N2	2	4.9	567	1456
N3	3	3.4	169	1010
N4	4	5.8	393	1723
N5	5	2.7	54	802
N6	6	1.5	11	446
N7	7	6.7	16	1991
N8	8	7.9	747	2348
N9	9	4.6	186	1367
N10	10	6.9	121	2050



**Figure 3.** Load and PV profile at node N10.

### 3.2. Reactive Power Demand in the Test Feeder

Figure 4 shows the voltage topology in the test feeder with zero reactive power feed-in for a static grid situation. As input data, the values from the load and PV profiles were used as listed in Table 2. The x- and the y-axis indicate the distance to the transformer. The transformer is circled in black, the households in gray. The fill color of each bus indicates its voltage.

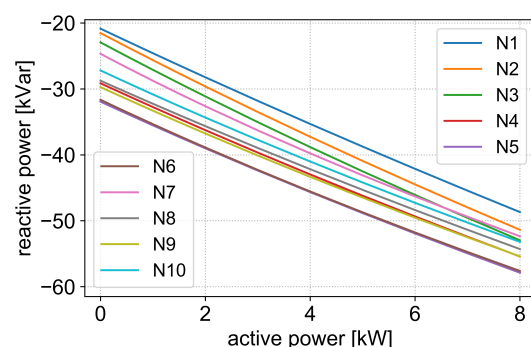


**Figure 4.** Voltage topology in the test feeder without reactive power injection.

At the transformer, the voltage is 1 pu and rises from there with the distance up to 1.02 pu due to the active power injections. Thus, Figure 4 illustrates the voltage rise along the lines in distribution grids with high PV penetration and demonstrates the risk for overvoltages at distant nodes. This emphasizes the need for voltage regulating measures, for instance additional reactive power injection.

The reactive power demand in the test feeder was analyzed. Based on the test feeder, it was investigated how much reactive power is necessary to regulate the voltage to 1 pu

under static PV feed-in. For this purpose, the reactive power at the grid connection point was systematically varied for different in-feeds with fixed active power until a voltage of 1 pu was observed. With exception of the node under consideration, all other nodes were kept constant according to Table 2. By this, the demand of reactive power was recorded as a function of the active power injection and is shown for every connection point in Figure 5.



**Figure 5.** Reactive power demand at nodes N1 to N10 to set local voltage to 1 pu.

The figure shows that the reactive power demand is individual for every connection point, which emphasizes the relevance of a self-learning volt-var control algorithm. In this case study, reactive power in the magnitude of  $-20$  kVar to  $-60$  kVar is required to maintain a voltage level of 1 pu depending on the active power feed-in. This comparatively high amount of reactive power results from the fact that a target voltage of exactly 1 pu was used in this study in order to show the theoretical potential of DRL VVC. However, in practice, most probably wider tolerance intervals will be allowed; thus, the reactive power demand reduces.

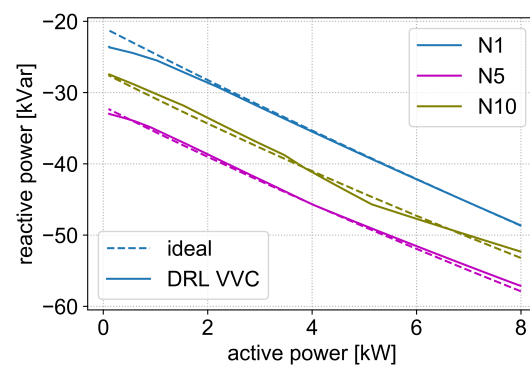
#### 4. Simulation Results—Application of the DRL Volt-Var Control Algorithm

##### 4.1. Static Grid Behavior

In this section, the performance of the proposed algorithm is investigated in a static grid environment. The power was kept constant at all nodes according to Table 2. At one node, the DRL agent was implemented and trained over a period of 80,000 steps. The aim was to control the voltage at 1 pu with an admissible control deviation of 0.2 %; thus, the parameter  $a$  in (1) was set to 0.002 pu. Nevertheless, other values are also possible for  $a$ , e.g., 0.05 pu in case less reactive power is available. During training, the active power feed-in at the considered node was varied according to the PV profile.

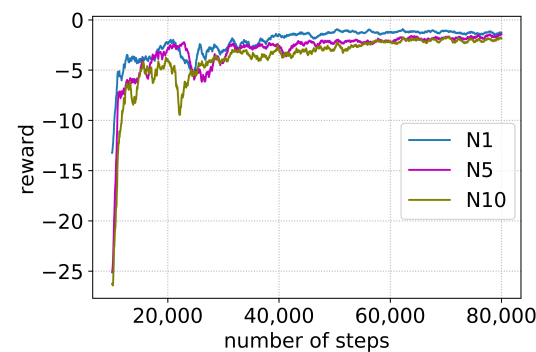
After the training process the active power was increased from 0 to 8 kW in 1000 steps to determine the characteristic curve for the reactive power output. The simulation has been carried out for the nodes N1, N5, and N10. As a result of these simulation runs, Figure 6 shows the reactive power over the active power. These characteristic curves can be interpreted as the reactive power curves learned by the DRL agent. These characteristic curves match very well with the corresponding ideal characteristic curves calculated in Section 3 and also shown in Figure 6.

The figure shows that the proposed algorithm has learned the individual reactive power demand in a static grid environment with high agreement to the theoretic benchmark. All learned curves show only slight deviations from the corresponding optimal curves. Minor deviations occur, especially in the range of very small and very large active power. This observation may be due to the fact that the training was carried out with real PV data instead of equally distributed data. Thus, some values are not presented during the training process. Despite the limited training data, the proposed algorithm was able to learn the reactive power demand for a large feed-in spectrum.



**Figure 6.** Reactive power output from the deep reinforcement learning (DRL) volt-var control (VVC) at the nodes N1, N5, and N10 after 80,000 training steps together with the calculated reactive power demand as a reference.

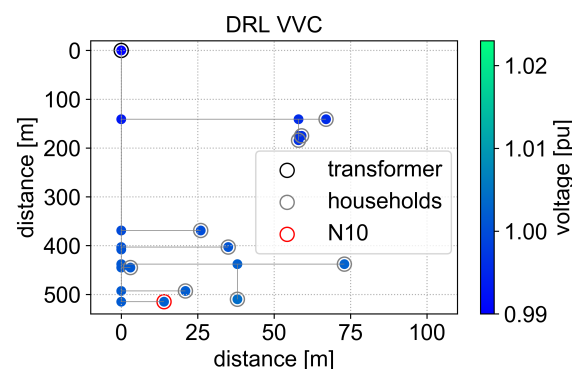
To visualize the training process, the moving average reward is shown in Figure 7. The moving average reward was calculated over 1000 successive learning steps at the three different nodes, N1, N5, and N10.



**Figure 7.** Moving average rewards over 1000 successive learning steps over the number of steps during static learning application of the proposed DRL VVC algorithm.

For all nodes, the proposed DRL algorithm shows an improvement of the reward with the number of training steps. This verifies the suitability of the presented reward function. After 60,000 steps, the moving average reward increases very slowly indicating a stable training result.

To investigate the effect of the DRL algorithm on adjacent nodes while applying the self-learning volt-var control at node N10 (red circle), the voltage at each bus was evaluated. The resulting voltage topology is presented in Figure 8. The x- and the y-axis indicate the distance to the transformer, whereas the fill color of each bus indicates its voltage.

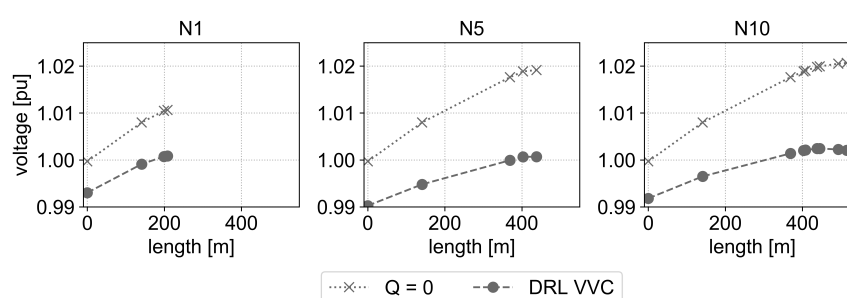


**Figure 8.** Voltage topology in the test feeder with reactive power injection at node N10 (red circle), the transformer is circled in black, the households in gray.



With the proposed DLR VVC algorithm the voltage at the transformer is about 0.992 pu and rises to 1.002 pu at distant nodes. It is noticeable that the voltage deviations in the network are significantly smaller than in Figure 4. For  $Q = 0$ , the voltage at the transformer was 1 pu and rises from there with the distance up to 1.02 pu due to the active power injections. Thus, the voltage difference when applying the proposed VVC is only half as large as in the reference with zero reactive power (Figure 4). Furthermore, with VVC, the voltage at the grid connection point of the smart inverter can be controlled within a 0.2 % interval of the nominal voltage. This value corresponds exactly to the tolerance range defined in the reward function ( $a = 0.002$  pu).

In addition to the voltage topology, the voltage at the nodes along the line from the transformer to the inverter was calculated. For the nodes N1, N5, and N10, the voltage rise along the line length  $l$  is shown in Figure 9 with and without self-learning volt-var control. The volt-var control was always applied at the node under consideration; all other nodes had no VVC.



**Figure 9.** Voltage along the line to nodes N1, N5, and N10 without reactive power feed-in and with DRL VVC at the considered node.

Without reactive power injection, the transformer voltage is 1 pu and rises up to 1.011, 1.019, or 1.021 pu, depending on the length of the line. The longer the line, the stronger is the voltage rise. Using the proposed volt-var control, the voltage at the respective node is lowered to 1 pu, with the result that the voltage along the line to the transformer continues to drop. The transformer voltage ranges between 0.99 and 0.993 pu for the different cases. In order to regulate the transformer voltage to  $1 \pm 0.002$  pu, as well, and thereby further reduce the voltage differences along the line, a combination of various smart inverters at different nodes could be investigated in future studies. Nevertheless, with the proposed DRL VVC algorithm the voltage differences along the line were reduced significantly by up to 50% in this case study.

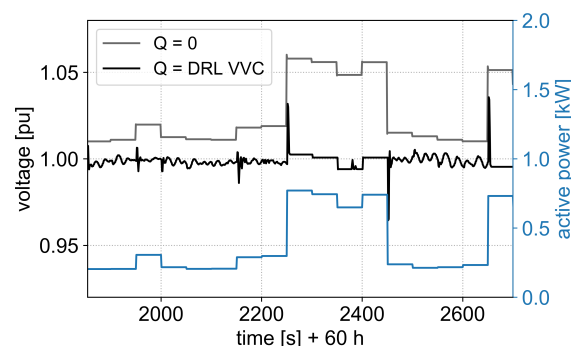
#### 4.2. Dynamic Grid Behavior

After successful application of the proposed DRL VVC algorithm in static grid environments, in this subsection the performance in a dynamic grid environment is investigated. For this purpose, the load and PV values were varied in time according to their profiles from Table 2. In all subsequent simulations, the time  $t = 0$  indicates 21 June, 00.00 h 00 s of the data set. As before, the data for phase A was used for all three phases. The active power feed-in changed every 50 s according to its PV profile. The test feeder was simulated over a period of 120 h (5 days), and the learning progress of the DRL agent was observed.

The self-learning volt-var control was located at node N10 as an online-learning algorithm without any additional knowledge or previous training. After only 3 simulated days of online training, the agent was able to balance the voltage at its connection point within the given interval; thus, the voltage rise at node N10 was eliminated. During the DRL training process the voltage deviations never exceeded the deviations of the reference without VVC. However, an inherent problem was observed. Every time the active power feed-in changes, a voltage peak occurs, due to the fact that the algorithm has a delay of one step because of the measurement duration. Thus, there is a time lag between the voltage change and the response of the algorithm in form of reactive power injection. Despite these



voltage peaks, a significant voltage reduction was achieved with the proposed DRL agent compared to the voltage curve without reactive power injection. Figure 10 shows a section of the voltage behavior with and without VVC at node N10.



**Figure 10.** Dynamic voltage behavior at node N10 with and without DRL VVC.

Ignoring the voltage peaks every 50 s, with DRL VVC a systematic voltage reduction of up to 0.06 pu can be observed at node N10. Under application of the volt-var control, the voltage at N10 ranges around 1 pu including small deviations because of the ongoing exploration by the DRL agent. These results verify the performance of the proposed DRL volt-var control algorithm and demonstrate its potential for future application in smart inverters.

## 5. Conclusions

This paper proposes a novel self-learning volt-var control algorithm on the basis of deep reinforcement learning. The algorithm is an online-learning DDPG that can be applied under real-time in smart inverters for reactive power management. In contrast to other machine learning-based volt-var control methods, no additional communication devices are needed. The only input data for the proposed algorithm are the measured voltage values at the grid connection point of the inverter.

The proposed DRL volt-var control was successfully tested in simulations at different nodes in a 21-bus low-voltage distribution grid. A significant voltage reduction was shown both in a static grid environment and a dynamic environment, and the proposed DRL algorithm was able to keep the voltage within the desired range of  $1 \pm 0.002$  pu. Furthermore, in the static case, the voltage difference along the lines were reduced by up to 50% using DRL VVC.

In this study, the aim was to control the voltage at 1 pu within a tolerance range of 0.2%. However, by adjusting the reward function, other intervals and use cases are also possible, such as higher tolerances or the most economical use of reactive power.

Thus, the proposed volt-var control algorithm is promising for application in real inverters and can make a significant contribution to decentralized, automated voltage control in future grids with high DG-penetration.

**Author Contributions:** Conceptualization, K.B., R.B., and S.G.; methodology, K.B. and R.B.; validation, R.B. and S.G.; investigation, K.B. and R.B.; writing—original draft preparation, K.B.; writing—review and editing, R.B. and S.G.; visualization, K.B.; supervision, K.v.M. and C.A.; project administration, S.G. and K.v.M.; funding acquisition, S.G., K.v.M., and C.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ueda, Y.; Kurokawa, K.; Tanabe, T.; Kitamura, K.; Sugihara, H. Analysis Results of Output Power Loss Due to the Grid Voltage Rise in Grid-Connected Photovoltaic Power Generation Systems. *IEEE Trans. Ind. Electron.* **2008**, *55*, 2744–2751. [\[CrossRef\]](#)
2. Tahir, M.; Nassar, M.E.; El-Shatshat, R.; Salama, M.M.A. A review of Volt/Var control techniques in passive and active power distribution networks. In Proceedings of the 2016 IEEE Smart Energy Grid Engineering (SEGE), Oshawa, ON, Canada, 21–24 August 2016; pp. 57–63. [\[CrossRef\]](#)
3. Woyte, A.; Thong, V.; Belmans, R.; Nijs, J. Voltage Fluctuations on Distribution Level Introduced by Photovoltaic Systems. *IEEE Trans. Energy Convers.* **2006**, *21*, 202–209. [\[CrossRef\]](#)
4. Chen, H.; Chen, J.; Shi, D.; Duan, X. Power flow study and voltage stability analysis for distribution systems with distributed generation. In Proceedings of the 2006 IEEE Power Engineering Society General Meeting, Montreal, QC, Canada, 18–22 June 2006; p. 8. [\[CrossRef\]](#)
5. Dong, F.; Chowdhury, B.; Crow, M.; Acar, L. Improving Voltage Stability by Reactive Power Reserve Management. *IEEE Trans. Power Syst.* **2005**, *20*, 338–345. [\[CrossRef\]](#)
6. Niknam, T.; Ranjbar, A.; Shinari, A. Impact of distributed generation on volt/var control in distribution networks. In Proceedings of the 2003 IEEE Bologna Power Tech Conference Proceedings, Bologna, Italy, 23–26 June 2003; Volume 3, pp. 210–216. [\[CrossRef\]](#)
7. Hietpas, S.; Naden, M. Automatic voltage regulator using an AC voltage-voltage converter. *IEEE Trans. Ind. Appl.* **2000**, *36*, 33–38. [\[CrossRef\]](#)
8. Chamana, M.; Chowdhury, B.H. Optimal Voltage Regulation of Distribution Networks With Cascaded Voltage Regulators in the Presence of High PV Penetration. *IEEE Trans. Sustain. Energy* **2018**, *9*, 1427–1436. [\[CrossRef\]](#)
9. Singh, B.; Solanki, J. A Comparison of Control Algorithms for DSTATCOM. *IEEE Trans. Ind. Electron.* **2009**, *56*, 2738–2745. [\[CrossRef\]](#)
10. Smith, J.W.; Sunderman, W.; Dugan, R.; Seal, B. Smart inverter volt/var control functions for high penetration of PV on distribution systems. In Proceedings of the 2011 IEEE/PES Power Systems Conference and Exposition, Phoenix, AZ, USA, 20–23 March 2011; pp. 1–6. [\[CrossRef\]](#)
11. Malekpour, A.R.; Pahwa, A. Reactive power and voltage control in distribution systems with photovoltaic generation. In Proceedings of the 2012 North American Power Symposium (NAPS), Champaign, IL, USA, 9–11 September 2012; pp. 1–6. [\[CrossRef\]](#)
12. Demirok, E.; González, P.C.; Frederiksen, K.H.B.; Sera, D.; Rodriguez, P.; Teodorescu, R. Local Reactive Power Control Methods for Overvoltage Prevention of Distributed Solar Inverters in Low-Voltage Grids. *IEEE J. Photovoltaics* **2011**, *1*, 174–182. [\[CrossRef\]](#)
13. Keane, A.; Ochoa, L.F.; Vittal, E.; Dent, C.J.; Harrison, G.P. Enhanced Utilization of Voltage Control Resources With Distributed Generation. *IEEE Trans. Power Syst.* **2011**, *26*, 252–260. [\[CrossRef\]](#)
14. Pereira, B.R.; Martins da Costa, G.R.M.; Contreras, J.; Mantovani, J.R.S. Optimal Distributed Generation and Reactive Power Allocation in Electrical Distribution Systems. *IEEE Trans. Sustain. Energy* **2016**, *7*, 975–984. [\[CrossRef\]](#)
15. Satsangi, S.; Kumbhar, G. Review on Volt/VAR Optimization and Control in Electric Distribution System. In Proceedings of the 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Delhi, India, 4–6 July 2016; pp. 1–6. [\[CrossRef\]](#)
16. Manbachi, M.; Farhangi, H.; Palizban, A.; Arzanpour, S. Smart grid adaptive volt-VAR optimization: Challenges for sustainable future grids. *Sustain. Cities Soc.* **2017**, *28*, 242–255. [\[CrossRef\]](#)
17. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M. Playing Atari with Deep Reinforcement Learning. *arXiv* **2013**, arXiv:1312.560.
18. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [\[CrossRef\]](#)
19. Kober, J.; Bagnell, J.A.; Peters, J. Reinforcement learning in robotics: A survey. *Int. J. Robot. Res.* **2013**, *32*, 1238–1274. [\[CrossRef\]](#)
20. Gu, S.; Holly, E.; Lillicrap, T.; Levine, S. Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 3389–3396. [\[CrossRef\]](#)
21. Cao, D.; Hu, W.; Zhao, J.; Zhang, G.; Zhang, B.; Liu, Z.; Chen, Z.; Blaabjerg, F. Reinforcement Learning and Its Applications in Modern Power and Energy Systems: A Review. *J. Mod. Power Syst. Clean Energy* **2020**, *8*, 1029–1042. [\[CrossRef\]](#)
22. Perera, A.; Kamalaruban, P. Applications of reinforcement learning in energy systems. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110618. [\[CrossRef\]](#)
23. Zhang, D.; Han, X.; Deng, C. Review on the research and practice of deep learning and reinforcement learning in smart grids. *CSEE J. Power Energy Syst.* **2018**, *4*, 362–370. [\[CrossRef\]](#)
24. Zhang, Z.; Zhang, D.; Qiu, R.C. Deep reinforcement learning for power system: An overview. *CSEE J. Power Energy Syst.* **2019**. [\[CrossRef\]](#)

25. Glavic, M. (Deep) Reinforcement learning for electric power system control and related problems: A short review and perspectives. *Annu. Rev. Control* **2019**, *48*, 22–35. [\[CrossRef\]](#)
26. Wang, W.; Yu, N.; Gao, Y.; Shi, J. Safe Off-Policy Deep Reinforcement Learning Algorithm for Volt-VAR Control in Power Distribution Systems. *IEEE Trans. Smart Grid* **2020**, *11*, 3008–3018. [\[CrossRef\]](#)
27. Yang, Q.; Wang, G.; Sadeghi, A.; Giannakis, G.B.; Sun, J. Two-Timescale Voltage Control in Distribution Grids Using Deep Reinforcement Learning. *IEEE Trans. Smart Grid* **2020**, *11*, 2313–2323. [\[CrossRef\]](#)
28. Li, C.; Jin, C.; Sharma, R. Coordination of PV Smart Inverters Using Deep Reinforcement Learning for Grid Voltage Regulation. In Proceedings of the 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; pp. 1930–1937. [\[CrossRef\]](#)
29. Zhang, Y.; Wang, X.; Wang, J.; Zhang, Y. Deep Reinforcement Learning Based Volt-VAR Optimization in Smart Distribution Systems. *IEEE Trans. Smart Grid* **2021**, *12*, 361–371. [\[CrossRef\]](#)
30. Cao, D.; Hu, W.; Zhao, J.; Huang, Q.; Chen, Z.; Blaabjerg, F. A Multi-Agent Deep Reinforcement Learning Based Voltage Regulation Using Coordinated PV Inverters. *IEEE Trans. Power Syst.* **2020**, *35*, 4120–4123. [\[CrossRef\]](#)
31. Wang, S.; Duan, J.; Shi, D.; Xu, C.; Li, H.; Diao, R.; Wang, Z. A Data-Driven Multi-Agent Autonomous Voltage Control Framework Using Deep Reinforcement Learning. *IEEE Trans. Power Syst.* **2020**, *35*, 4644–4654. [\[CrossRef\]](#)
32. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous control with deep reinforcement learning. *arXiv* **2015**, arXiv:1509.02971.
33. Liu, H.; Wu, W. Two-stage Deep Reinforcement Learning for Inverter-based Volt-VAR Control in Active Distribution Networks. *IEEE Trans. Smart Grid* **2020**. [\[CrossRef\]](#)
34. Silver, D.; Lever, G.; Heess, N.; Degris, T.; Wierstra, D.; Riedmiller, M. Deterministic policy gradient algorithms. In Proceedings of the 31st International Conference on Machine Learning (ICML 2014), Beijing, China, 21–26 June 2014; Volume 1, pp. 605–619.
35. Brockman, G.; Cheung, V.; Pettersson, L.; Schneider, J.; Schulman, J.; Tang, J.; Zaremba, W. OpenAI Gym. *arXiv* **2016**, arXiv:1606.01540v1.
36. Plappert, M. keras-rl. *GitHub Repos.* **2016**. Available online: <https://github.com/keras-rl/keras-rl> (accessed on 19 February 2021).
37. Köppl, S.; Bruckmeier, A.; Böning, F.; Hinterstocker, M.; Kleinertz, B.; Konetschny, C.; Mueller, M.; Samweber, F.; Schmid, T.; Zeiselmaier, A. *Projekt MONA 2030: Grundlage für die Bewertung von Netzoptimierenden Maßnahmen: Teilbericht Basisdaten*; Forschungsstelle für Energiewirtschaft e.V. (FfE): München, Germany, 2017.
38. Dassault Systèmes. FMKit for Simulink. **2020**. Available online: <https://github.com/CATIA-Systems/FMKit-Simulink> (accessed on 19 February 2021).
39. Tjaden, T.; Joseph, B.; Weniger, J.; Quaschnig, V. *Repräsentative Elektrische Lastprofile für Einfamilienhäuser in Deutschland auf 1-Sekündiger Datenbasis*; Technical Report; Hochschule für Technik und Wirtschaft Berlin: Berlin, Germany, 2015. [\[CrossRef\]](#)
40. Weniger, J.; Quaschnig, V. Begrenzung der Einspeiseleistung von netzgekoppelten Photovoltaiksystemen mit Batteriespeichern. In Proceedings of the 28. Symposium Photovoltaische Solarenergie, Staffelstein, Germany, 6–8 March 2013.